

Analyzing the Relationships of Psychological Symptoms in the SCL-90 Using Genetic Algorithms and Association Rule Mining

Qiang Wang *, Daisy S. Yap

Graduate School, Angeles University Foundation, Angeles City, Philippines

*Corresponding Author: Qiang Wang (Email: wang.quiang@auf.edu.ph)

ABSTRACT

To improve the efficiency of processing SCL-90 data with association rules, this study integrates genetic algorithms with association rule mining techniques to analyze data from the SCL-90. The goal is to identify potential patterns and associations in psychological health issues. The SCL-90 is a widely used tool for assessing psychological symptoms and distress levels, consisting of 90 items that cover multiple dimensions such as obsessive-compulsive symptoms, anxiety, and depression. By employing genetic algorithms for preprocessing and optimizing the SCL-90 data, representative and distinctive feature subsets are selected. Genetic algorithms simulate the processes of natural selection, crossover, and mutation to identify optimal feature combinations, thereby enhancing the efficiency and effectiveness of association rule mining. The selected data is then analyzed using association rule mining techniques to generate rules with high support and confidence, revealing potential links between different psychological symptoms. The combination of genetic algorithms and association rule mining not only improves data processing and pattern discovery efficiency but also provides valuable insights for psychological health assessment and intervention.

KEYWORDS

Association Rule; SCL90; Genetic Algorithm; Machine Learning

1. INTRODUCTION

In today's society, mental health issues are a growing concern as the pace of life and competitive pressures increase. Globally, the incidence of mental health problems has risen significantly, with depression, anxiety and obsessive-compulsive disorders becoming more common. The World Health Organization (WHO) reports that about 300 million people worldwide suffer from depression, and the number of people with anxiety disorders is also increasing year by year. These psychological problems not only affect the quality of life of individuals, but also impose a huge burden on families and society. For this reason, how to effectively assess, understand and intervene in mental health problems has become a major challenge that needs to be addressed.

In mental health research and clinical diagnosis, it is crucial to accurately assess an individual's psychological symptoms and level of distress. The Symptom Self-Control Scale (SCL-90) is a widely used psychological assessment tool that comprehensively covers multiple dimensions of mental health through 90 items, including obsessive-compulsive symptoms, anxiety, depression, hostility, somatization, interpersonal sensitivity, paranoia, psychoticism, and phobia. These items help psychologists and clinicians identify and quantify a patient's psychological symptoms and provide a scientific basis for developing treatment plans. However, with the increasing volume of data and the complexity of mental health problems, traditional analytic methods face many challenges.

The development of modern data mining techniques provides new tools and methods for mental health research. Data mining techniques can extract useful information and patterns from large amounts of complex data, revealing potential associations and trends. However, data mining techniques also face certain difficulties when dealing with high-dimensional data such as SCL-90. In particular, association rule mining may face the problem of dimensional catastrophe when dealing with high-dimensional data, i.e., the increase in data dimensionality leads to a sharp increase in computational complexity and storage requirements in order to effectively solve this problem, this paper proposes an approach that combines Genetic Algorithm and Association Rule Mining [1]. Genetic algorithm is an optimization algorithm based on natural selection and genetic mechanism, which is suitable for complex search space and multi-objective optimization problems. By simulating selection, crossover and mutation in the process of biological evolution, genetic algorithms can efficiently find near-optimal solutions in a large search space [2].

The main purpose of applying genetic algorithm to SCL-90 data analysis is to reduce the data dimensions and improve the efficiency and effectiveness of association rule mining through feature selection and optimization.

In practice, the genetic algorithm first generates an initial population by randomly generating an initial population in which each individual represents a combination of features in the SCL-90 data. Then, the strengths and weaknesses of each individual are evaluated by the fitness function, and the individual with higher fitness is selected as the parent to perform crossover and mutation operations to generate new offspring. This process is continuously iterated to finally find the feature combination with the best fitness. With this approach, we can filter out the most representative and differentiated features and significantly reduce the data dimensionality

After feature selection and optimization, we use association rule mining techniques to analyse the filtered data. Association rule mining aims to discover the relationship between frequent itemsets in a dataset and measures the importance and reliability of rules through support and confidence. Support indicates how often a rule occurs in a data set and confidence indicates the reliability of the rule. In SCL-90 data analysis, potential links between different psychological symptoms can be revealed through association rule mining [3]. For example, the co-occurrence of depressive symptoms and anxiety symptoms, or the association between somatization symptoms and interpersonal sensitivity, these rules are important for understanding the co-occurrence mechanism of psychological symptoms.

By combining genetic algorithms with association rule mining, significant results were achieved in processing high-dimensional SCL-90 data. The genetic algorithm effectively performs feature selection and optimization, mitigates the dimensionality catastrophe problem faced by association rule mining, and improves the efficiency and effectiveness of rule mining. The final analysis results provide valuable references for mental health professionals to better understand the associations between different psychological symptoms and provide a scientific basis for the development of personalized treatment and intervention strategies.

In this study, SCL-90 data were analysed in depth by combining genetic algorithms with association rule mining techniques to reveal potential patterns and associations of mental health problems. This method is not only innovative in theory, but also shows good application prospects in practice, providing new tools and methods for mental health research and clinical practice. In the future, we will further optimize the algorithm and explore more application scenarios to promote the development and application of mental health data mining technology.

2. RELATED WORK

2.1. Literature Review

The SCL-90 (Symptom Checklist-90) is a widely used tool for assessing individual psychological symptoms, covering nine dimensions such as somatization, obsessive-compulsive symptoms,

interpersonal sensitivity, depression, anxiety, hostility, phobia, paranoia, psychoticism and others. With the development of data mining technology, how to effectively analyze the psychological symptom relationships in SCL-90 and reveal the underlying patterns and associations has become a hot topic for researchers. Genetic algorithm and association rule analysis, as two powerful data mining tools, have shown great potential for application in psychological research [4]. The SCL-90 is designed to assess an individual's psychological symptoms over a recent period of time through a self-rating scale. It consists of 90 items, each of which reflects the severity of an individual's symptoms through a five-point scale [5]. The SCL-90 is widely used in the fields of clinical psychology, psychiatry, and psychotherapy, and is an important tool in the study of mental health problems. Genetic Algorithm (GA) is an optimization algorithm based on natural selection and genetic mechanism. By simulating the process of biological evolution, Genetic Algorithm shows its unique advantages in solving complex optimization problems [6]. Its basic steps include selection, crossover, mutation and fitness assessment. Examples of applications of genetic algorithms in psychological research include optimization of psychometric instruments and identification of psychological symptom patterns [7]. Association rule algorithms are a technique used to discover relationships between items in a data set. By calculating support, confidence, and lift, association rule analysis can reveal potential associations between data items [8]. In psychology and medical research, association rule analysis is used to mine the relationships between data such as psychological symptoms, disease symptoms, etc., providing new research perspectives and methods. Combining genetic algorithms with association rule analysis allows for more effective pattern discovery on large datasets. Genetic algorithms can optimize the association rule mining process, improving the accuracy of the rules and discovering potentially complex patterns [9]. It has been shown that combining these two approaches can improve the effectiveness of data analysis in a variety of fields, such as market analysis and medical diagnosis [10]. In specific implementation, the SCL-90 data can be initially processed using genetic algorithms to select important symptom dimensions and items, and then the potential relationship between these symptoms can be mined through association rule analysis [11]. By way of example studies, associations between, for example, depression and anxiety, or the relationship between certain somatization symptoms and psychological symptoms can be discovered, thereby providing a scientific basis for mental health assessment and intervention. Although combining genetic algorithms and association rule analysis shows great potential in SCL-90 data analysis, it also faces some challenges, such as data quality issues and computational complexity. Future research can focus on improving the algorithm, enhancing computational efficiency, and exploring more application scenarios, such as real-time mental health monitoring and personalized psychological intervention. Combining genetic algorithm and association rule analysis provides a new approach to the study of SCL-90 psychological symptom relationships, which is innovative and important. A systematic literature review can provide theoretical support and practical guidance for this research direction, and provide new ideas and methods for mental health assessment and intervention.

2.2. Methodology

2.2.1. Data source

The data were sourced from the SCL-90 survey completed by new students at a particular university in September 2021. A total of 3,278 individuals participated, resulting in 3,278 sets of data. The results are presented in the following table 1:

Table 1. Preprocessed data structures

ID	S	O	I	D	A	H	T	U	P	E
1	1	1	1	1	1	1	1	2	1	1
2	1	1	1	1	1	1	1	1	1	1
...

S=Degree of Somatization, O=Degree of Obsessive-Compulsive Symptoms, I=Degree of Interpersonal Relationships, D= Degree of Depression, A= Degree of Anxiety, H= Degree of Hostile, T= Degree of Terror, U= Degree of Stubborn, P= Degree of Psychotic, E= Degree of Other. Each of these 10 psychological states is divided into 5 ratings based on a Likert scale. The term “impact” encompasses both the distress and anxiety elicited by the symptoms, as well as the disruption to psychosocial functioning attributable to these symptoms. The definitions of "mild," "moderate," and "severe" are subjective and are based on the self-evaluator's personal experience, with no strict criteria imposed. For the purposes of the subsequent experimental procedures, the five response categories in this SCL-90 survey are quantified as follows: 1 for none, 2 for mild, 3 for moderate, 4 for severe, and 5 for extremely severe.

2.2.2. Implementation of the genetic algorithm

Design an appropriate way to encode individuals in order to represent candidate solutions in the genetic algorithm. Then define a fitness function that measures the degree of superiority or inferiority of each individual in the solution space to guide the evolutionary process [12]. Selects the individual with higher fitness as the parent for reproducing the next generation. Design crossover and mutation operations for generating new individuals and maintaining the diversity of the population. Perform selection, crossover and mutation operations iteratively until the stopping condition is satisfied.

2.2.3. Association rule analysis

Utilize association rule algorithms, such as Apriori algorithm, to mine frequent item sets from SCL-90 data. Generate association rules with certain confidence and support based on frequent itemset [13].

Evaluate the quality of the generated rules and filter out the rules with high correlation with psychological symptoms.

The computational results of association rules are judged by support, confidence, and lift [14].

Support:

Support measures how often a particular item set appears in the dataset. Specifically, for association rules $X \rightarrow Y$

$$\text{Support}(X \rightarrow Y) = \frac{\text{transactions containing } X \text{ and } Y}{\text{total transactions}} \quad (1)$$

The support indicates the simultaneous inclusion of the itemset in all transactions X and Y as a percentage of transactions. High support indicates that the rule $X \rightarrow Y$ is common in the dataset.

Confidence:

Confidence measures the proportion of transactions containing itemset X that also contain itemset Y . For an association rule $X \rightarrow Y$, the confidence is defined as:

$$\text{Confidence}(X \rightarrow Y) = \frac{\text{Support}(X \rightarrow Y)}{\text{Support}(X)} \quad (2)$$

Confidence indicates the likelihood of itemset Y appearing in transactions that contain itemset X . A high confidence value means itemset Y is likely to appear in transactions that contain itemset X .

Lift:

Lift measures the independence of itemsets X and Y . For an association rule $X \rightarrow Y$, the lift is defined as:

$$\text{Lift}(X \rightarrow Y) = \frac{\text{Support}(X \rightarrow Y)}{\text{Support}(X) \times \text{Support}(Y)} \quad (3)$$

A lift value greater than 1 indicates a positive correlation between itemsets X and Y , meaning the occurrence of X increases the likelihood of Y occurring. A lift value equal to 1 indicates independence, and a lift value less than 1 indicates a negative correlation.

2.2.4. Combination of genetic algorithms and association rules

Combine individual coding in genetic algorithm with association rule representation to design appropriate coding scheme. Then, incorporate the indexes of support and confidence of association rules into the fitness function to evaluate the advantages and disadvantages of candidate solutions. Search the solution space using genetic algorithm to find the optimal combination of association rules to reveal the potential relationship between psychological symptoms.

2.2.5. Experimental design and evaluation

Determine the experimental parameters, such as population size, number of iterations, crossover probability, mutation probability, etc. Assess the validity and interpretability of the obtained association rule combinations in describing psychological symptom relationships. Analyze the association rules found to explain the correlations between different psychological symptoms and provide new insights for mental health research [15].

2.2.6. Strengths and limitations of the method

Point out the advantages of the method over traditional single-analysis methods, such as the ability to find more complex psychological symptom relationships. Explain the possible limitations of the method, such as the sensitivity to parameters and the computational complexity.

3. RESULT

In the context of the same dataset, in the first experiment the results were obtained through the association rule algorithm, and in the second experiment the dataset was first optimized using the genetic algorithm, and then the association rule algorithm was used to obtain the results, and it should be said in particular that in the second experiment the genetic algorithm will use the gridded automated optimization process in optimizing the data, which has the advantage of running automatically after setting the range, and there is no need to manually tune the parameters to seek the optimal solution. The advantage of this is that it will run automatically after the range is set and does not need to be manually adjusted to seek the optimal solution.

In the association rule the support will be set to 0.2 in order to produce frequent terms greater than 0.2 [16]. The population size, iteration number, crossover probability and mutation probability will be set to [50,100], [50,80], [0.6,0.7], [0.01,0.05] in the genetic algorithm, which is conducive to the optimal solution obtained by the grid-based automated operation [17].

The following experimental results will be obtained by the means mentioned above.

The results, calculated using Python's association rule algorithm, are shown in the following table 2 and table 3:

Table 2. Frequent Itemsets

	Support	Itemsets
0	0.38	S
1	0.85	O
2	0.34	I
3	0.34	D
4	0.35	(S, O)
5	0.27	(I, O)
6	0.29	(D, O)

Table 3. Association Rules

antecedents	consequent	antecedent support	consequent support	support
S	O	0.38	0.85	0.35
I	O	0.34	0.85	0.27
D	O	0.34	0.85	0.29
confidence	lift	leverage	conviction	
0.92	1.08	0.03	1.91	
0.79	0.93	-0.02	0.71	
0.87	1.03	0.00	1.18	

The association rule algorithm produces six frequent items, which are S with support 0.38, O with support 0.85, I with support 0.34, D with support 0.34, (S, O) with support 0.35, confidence 0.92, and lift 1.08. (I, O) with support 0.27, confidence 0.79, lift 0.93. (D, O) with support 0.29, confidence 0.87, and lift 1.03.

Combining the genetic algorithm into the association rule algorithm resulted in the results shown in the table 4 and table 5 below:

Table 4. Frequent Itemsets

	Support	Itemsets
0	0.38	S
1	0.85	O
4	0.35	(S, O)

Table 5. Association Rules

antecedents	consequent	antecedent support	consequent support	support
S	O	0.38	0.85	0.35
confidence	lift	leverage	conviction	
0.92	1.08	0.03	1.91	

The association rule algorithm produces six frequent items, which are S with support 0.38, O with support 0.85, (S, O) with support 0.35, confidence 0.92, and lift 1.08.

Association rule mining optimized by genetic algorithm can effectively reduce redundant data and improve the quality and practical application value of rules. In scenarios where the data volume is large and manual screening is required, this optimization method can significantly improve the work efficiency and help staff find valuable information and make business decisions faster.

4. CONCLUSION

In this study, we improved the efficiency of SCL-90 data processing and identified potential patterns and associations in mental health problems by combining genetic algorithms with association rule mining techniques. SCL-90, as a widely used psychological symptom assessment tool, covers multiple dimensions such as obsessive-compulsive symptoms, anxiety, depression, etc., through 90 items. Genetic algorithms were used for preprocessing and optimization to select a subset of representative and unique characteristics. The genetic algorithm identifies the optimal combination of features by simulating the process of natural selection, crossover, and mutation, which improves the efficiency and effectiveness of association rule mining. Association rule mining on selected data

generates rules with high support and confidence, revealing potential links between different psychological symptoms. The combination of genetic algorithm and association rule mining not only improves the efficiency of data processing and pattern discovery, but also provides valuable insights for mental health assessment and intervention.

REFERENCES

- [1] J. Wu, "Evaluation Model of Product Shape Design Scheme Based on Fuzzy Genetic Algorithm Mining Spatial Association Rules," *Math. Probl. Eng.*, vol. 2022, pp. 1–10, Mar. 2022, doi: 10.1155/2022/2888472.
- [2] S. abd alwahab, "Genetic based method for Mining Association Rules from Text," 2021, doi: 10.17762/turcomat.v12i9.4795.
- [3] B. Xu, S. Ding, and Y. Li, "Data association rules mining method based on genetic optimization algorithm," *J. Phys.: Conf. Ser.*, vol. 1570, p. 012006, Jun. 2020, doi: 10.1088/1742-6596/1570/1/012006.
- [4] X. Liu, "Effect of a Mindfulness-Based Intervention Program on Comprehensive Mental Health Problems of Chinese Undergraduates," *Community Ment Health J*, vol. 55, no. 7, pp. 1179–1185, Jun. 2019, doi: 10.1007/s10597-019-00426-4.
- [5] S. Casale, S. E. Caplan, and G. Fioravanti, "Positive metacognitions about Internet use: The mediating role in the relationship between emotional dysregulation and problematic use," *Addict. Behav.*, vol. 59, pp. 84–88, Aug. 2016, doi: 10.1016/j.addbeh.2016.03.014.
- [6] M. Li et al., "Psychometric Properties and Measurement Invariance of the Brief Symptom Inventory-18 Among Chinese Insurance Employees," *Front. Psychol.*, vol. 9, Apr. 2018, doi: 10.3389/fpsyg.2018.00519.
- [7] G. H. Franke, S. Jaeger, H. Glaesmer, C. Barkmann, K. Petrowski, and E. Braehler, "Psychometric analysis of the brief symptom inventory 18 (BSI-18) in a representative German sample," *BMC Med Res Methodol*, vol. 17, no. 1, Jan. 2017, doi: 10.1186/s12874-016-0283-3.
- [8] B. Ghazanfari, F. Afghah, and M. E. Taylor, "Sequential Association Rule Mining for Autonomously Extracting Hierarchical Task Structures in Reinforcement Learning," *IEEE Access*, vol. 8, pp. 11782–11799, 2020, doi: 10.1109/access.2020.2965930.
- [9] B. Minaei-Bidgoli, R. Barmaki, and M. Nasiri, "Mining numerical association rules via multi-objective genetic algorithms," *Inf. Sci.*, vol. 233, pp. 15–24, Jun. 2013, doi: 10.1016/j.ins.2013.01.028.
- [10] N. Khuda Bux, M. Lu, J. Wang, S. Hussain, and Y. Aljeroudi, "Efficient Association Rules Hiding Using Genetic Algorithms," *Symmetry*, vol. 10, no. 11, p. 576, Nov. 2018, doi: 10.3390/sym10110576.
- [11] F. Peng, Y. Sun, Z. Chen, and J. Gao, "An Improved Apriori Algorithm for Association Rule Mining in Employability Analysis," *Teh. vjesn.*, vol. 30, no. 5, Oct. 2023, doi: 10.17559/tv-20230327000481.
- [12] V. Javangula, "Mining of High Utility Item sets using Genetic Algorithm," 2021, doi: 10.17762/turcomat.v12i9.3724.
- [13] J. Lei, "Association Rule Mining Algorithm in College Students' Quality Evaluation System," *J. Electr. Comput. Eng.*, vol. 2022, pp. 1–9, Apr. 2022, doi: 10.1155/2022/6721504.
- [14] U. M. Haque, E. Kabir, and R. Khanam, "Investigating school absenteeism and refusal among Australian children and adolescents using Apriori association rule mining," *Sci Rep*, vol. 14, no. 1, Jan. 2024, doi: 10.1038/s41598-024-51230-4.
- [15] F. Petry and R. Yager, "Data Mining Using Association Rules for Intuitionistic Fuzzy Data," *Information*, vol. 14, no. 7, p. 372, Jun. 2023, doi: 10.3390/info14070372.
- [16] R. Agrawal, T. Imieliński, and A. Swami, "Mining association rules between sets of items in large databases," in *Proceedings of the 1993 ACM SIGMOD international conference on Management of data - SIGMOD '93*, New York, New York, USA: ACM Press, 1993, pp. 207–216. doi: 10.1145/170035.170072.
- [17] D. Goldberg, "Genetic algorithms in search, optimization, and machine learning," *Choice Rev. Online*, vol. 27, no. 02, pp. 27-0936-27-0936, Oct. 1989, doi: 10.5860/choice.27-0936.